8. Data Visualization

Charles Kostelnick
Iowa State University

Data visualization is generally defined as the graphical representation of quantitative data related to virtually any subject or discipline—from business, engineering, and ***science*** to medicine, social science, mathematics, and statistics. Data displays appear in a wide range of communications: technical reports, journal articles, PowerPoint presentations, feasibility studies, annual reports, newsletters, and popular media, as well as in everyday documents like energy bills, investments statements, dashboards, and health and fitness records.

Data ***design*** can be classified into three basic plotting systems: rectilinear grids (bar charts, line graphs, scatterplots), circular configurations (pie charts, polar charts), and maps (Yau, 2013, pp. 104-107), though other classification systems have been proposed (Desnoyers, 2011). Robert Harris (1996) provided a detailed compendium of the hundreds of display ***genres*** used in the late 20th century, and since then, with the emergence of digital and interactive designs, many new and hybrid forms have begun to appear. However, traditional genres such as bar charts, line graphs, pie charts, maps, and scatterplots continue to be among the most popular forms of data design. Several of the genres commonly used today appear in online compendia (Eppler & Muntwiler, 2022; Ferdio ApS, 2021; Ribecca, 2023).

The ***history*** of data design has unfolded primarily in the past 300-400 years. Although rare earlier examples exist, graphical displays emerged in the 17th and 18th centuries as a means to chart weather and other scientific data and eventually to visualize engineering data. In the late 18th century, William Playfair (1801) applied graphical techniques to chart economic data, and during the so-called "golden age" of data design (Friendly, 2008; Funkhouser, 1937, p. 330) in the second half of the 19th century, visualizing data about population, health, and other human subjects developed rapidly, along with new forms, techniques, and genres. These developments coalesced with the creation of national atlases, most notably in the US and France, that visualized statistical data about nation states (see Kostelnick, 2004). Figure 8.1 shows a series of charts (mosaics) from the first *Statistical Atlas of the United States* (Walker & U.S. Census Office, 1874) that use rectilinear areas to show the relative populations of states and territories, arranged from smallest to largest.

With the emergence of modernism in the early 20th century, charts and graphs were simplified to create stronger ***visual*** impact and to appeal to larger ***public audiences*** (see Sloane & U.S. Bureau of the Census, 1914). At the same time, as data design began to establish a global presence, many additional discipline-specific

DOI: https://doi.org/10.37514/TPC-B.2023.1923.2.08

forms began to appear to show scientific, engineering, business, and medical data. In the later 20th century, digital tools enabled the rapid proliferation of data design, allowing anyone with the basic software to generate charts and graphs.

For most of data design's ***history***, charts and graphs have appeared in static paper form, whereby audiences interpret one fixed and immovable version. However, with the advent of interactive digital design, charts displaying quantitative data began to give users greater control by enabling them to choose which data to visualize, to modify the graphical display (genre, color), and to mouse over data points for additional details. This kind of display radically differs from static designs by giving users greater agency (Rawlins & Wilson, 2014).

The history of the early development of data design has been documented by H. Gray Funkhouser (1937), and the key graphical inventions and the pioneer designers who created them have been charted by Michael Friendly and Daniel Denis (2001-2018) in their website *Milestones in the History of Thematic Cartography, Statistical Graphics, and Data Visualization*. In addition, historical studies of genre, science, statistics, and mapping appear in *Visible Numbers: Essays on the History of Statistical Graphics* (Kimball & Kostelnick, 2016).

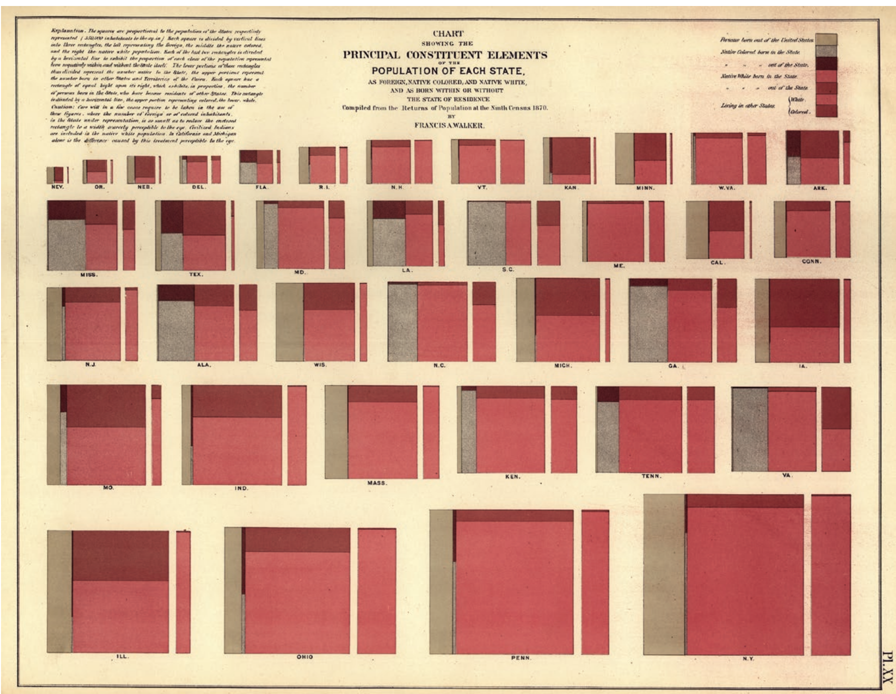

Figure 8.1. Mosaic charts from the Statistical Atlas of the United States *showing the composition of the population of states and territories (Walker & U.S. Census Office, 1874, Plate XX). Courtesy of the Library of Congress, Geography and Map Division.*

Data visualization has many contested areas, beginning with defining the kinds of images that actually fall within its realm, which can vary from one theorist or practitioner to another. Because data visualization can include virtually any image that represents "data" about some phenomenon, defining what fits within its boundaries can be a bit murky. Some visuals like pictures are usually not considered within the realm of data visualization because they typically don't represent quantitative data, though even pictures can be used for this purpose, especially when aligned in a series for comparison. Also typically outside the realm of data visualization are textual displays like diagrams, infographics, and organizational charts, though tables sometimes appear adjacent to graphical displays to provide precise numerical values. However, data visualization primarily refers to the graphical display of quantitative data.

Another area of debate concerns whether data visualization should be guided by perceptual principles or ***rhetorical*** factors. Many theorists and practitioners focus on perceptual efficiency: what forms audiences (or users) can most easily and efficiently process. For example, perceptual issues include the benefits and drawbacks of certain forms of graphical coding (lines, angles, areas, volume, gradients), which have been measured empirically (Cleveland & McGill, 1984; Heer & Bostock, 2010; see also Cochran et al., 1989; Macdonald-Ross, 1977). Other perceptual issues include "preattentive processing" (what users see in an instant), color recognition, and Gestalt principles (Few, 2012, pp. 66-68, 75-86). In the realm of digital media, perceptual issues include the effects of interactivity and animation (Fisher, 2010) on users' visual processing.

From a rhetorical perspective, these perceptual elements affect clarity, which is critical to any effective display; however, other aspects of audiences' interpretations also matter rhetorically: genres and their ability to meet audience expectations (Kostelnick, 2016a), ***ethical*** issues created by distortions or by lack of empathy (Dragga & Voss, 2001), emotional responses aroused by color or other graphical cues (Kostelnick, 2016b), agency afforded by interactivity (Rawlins & Wilson, 2014), and cultural and ideological influences on designers and on their audiences' interpretations (Barton & Barton, 1993a; Battle-Baptiste & Rusert, 2018; Brasseur, 2003; Li, 2020).

The perceptual and rhetorical approaches, however, need not be viewed as conflicting with each other. Indeed, they can complement each other as well, with the perceptual approach providing universal guidelines that guide functional communication and the rhetorical approach enabling designers to adapt their displays to specific audiences and situations.

Another distinction is often made between presentation charts, which are intended to inform or persuade audiences, and analytical charts, which are used as discovery tools to find hidden patterns in the data (see Fisher, 2010, pp. 338-339). Presentation charts are simpler and more explicit, while analytical charts are more complex, creative, and provisional. This dichotomy makes an important distinction about the data designer's processes and intentions: On the one hand, the designer carefully controls the data and the graphical elements in the display to ensure a good fit with audience and purpose, and on the other, the designer tries to model data with whatever tools and ***technology*** are available, with mixed and unpredictable results.

However, distinguishing between these two types of data design can be problematic, given the variations among audiences. A presentation chart that might be direct and compelling to one audience might be elementary or redundant to another; conversely, an analytical chart that's confusing and incomprehensible to one audience might be transparent and engaging to another. Nonetheless, the differences between these two modes of design are conceptually and operationally valid, with one emphasizing communication and the other emphasizing discovery.

Data designers often differ in their views about embellishing charts and graphs. Overly embellished charts, labeled "chartjunk" by Edward Tufte (1983, pp. 107-121), can impede clarity, and many data designers advise against excessive ornament. Modernist aesthetics also reinforced the minimalist approach to design, epitomized in the abstract pictographic system of Otto Neurath (1939). However, charts designed for popular media (e.g., those of Nigel Holmes, 1984) often contain illustrations and other embellishments to signal the subject of a given display and reveal its primary message. Moreover, with the advent of digital design, the uses of color and other non-data graphical elements have increased. So on the one hand, the purists prefer lean charts that reveal plentiful data as transparently as possible, while the artists see chart design as a way of engaging audiences, often those unmotivated to explore data. Both of these approaches have their place in data visualization, and the creativity of contemporary interactive design often bridges the two.

Ethical issues have long pervaded data design, though theorists and practitioners have defined them from several perspectives. Data design in popular media has been especially scrutinized because of deceptive practices, some of which Tufte (1983) demonstrates with his "Lie Factor," whereby perspective, volume, and area are misused (pp. 53-77; see also Brinton, 1914, pp. 20-27). Other examples of what are considered flawed (and potentially unethical) practices include starting the Y-axis scale above zero, stretching the plot frame vertically or horizontally to emphasize (or de-emphasize) trends or relationships, and using a double scale on the left and right sides of the plot frame. However, these methods might also be used ethically depending on the audience and situational context for a given display.

Data design can be evaluated according to general ethical principles (Kienzler, 1997), in relation to power and gender (D'Ignazio & Klein, 2020), as well as on its ability to project empathy and emotion (Dragga & Voss, 2001). However, the intentions of designers rarely matter, as designers are held accountable for displays that mislead their audiences. Still, ethical standards for data design vary, even for communications like annual reports, financial statements, and risk assessments, where charts and graphs might influence the audience's decision-making.

Data displays enable audiences to perceive the big patterns as well as explore the details, to see both the forest and the trees. Tufte (1990, pp. 37-51) described these two viewpoints as the "macro" and "micro" levels, and Ben Barton and Marthalee Barton (1993b) analyzed the differences between the "synoptic" and "analytic" ways of viewing. The macro-level (synoptic) view allows audiences to see the data patterns *at a glance*. As Playfair (1801) claimed early on in his pioneering work, charts and graphs enable us to perceive data "under one simple impression of vision," which makes them superior to tabular displays of data (p. x). Along similar lines, Jacques Bertin (1981) demonstrates the perceptual power and immediacy of our ability to "SEE" data graphically as opposed to interpreting data piecemeal (pp. 178-179).

Although the macro-level, big-picture view has always been touted as the most beneficial in visualizing data, ideally a chart should also allow users to explore data in detail and with precision at a more deliberate pace. Balancing these two levels of viewing creates challenges for designers, especially those working in print: gridlines, data labels, and more minute graphical coding can compromise macro-level processing, and space limitations can curb micro-level viewing. Interactive charts can address these problems by providing access to both levels through multiple viewing options. Because interactive charts afford user control over which data to display, they typically allow for both macro- and micro-level viewing.

Figure 8.2 shows an interactive animated scatterplot that uses both the macro- and micro-levels to visualize the relationship between life expectancy and income in countries around the world (Rosling et al., 2021). Users can see the big picture by viewing the animation of all countries over a span of over 200 years, or they can explore the details by stopping at a given year or by selecting a specific country from the menu on the right. Although creating interactive displays has heretofore been confined to a relatively small number of designers, software programs like Tableau (2023) are making interactive data visualization increasingly accessible.

The sprawling domain of data visualization is rapidly evolving and expanding as it shifts from print to digital forms, integrating the old and the new, generating hybrid forms, and sometimes reviving past forms. Currently, these developments are reshaping audiences' interactions with charts and graphs. In the future, digital design will likely generate novel and creative forms that will enable audiences to explore large data sets in stimulating and productive ways. At the same time, these new and inventive forms will challenge audiences perceptually and rhetorically, as audiences may have to experience a learning curve as they try to interpret them. However, audiences will be richly rewarded for their patience and interpretive flexibility. Digital design will also become increasingly dynamic in the future, as fluid data sets will be constantly replenished, giving audiences continuous visual access to the numbers. The data sets underlying these visualizations may be raw and unfiltered, but they will also be supple and timely.

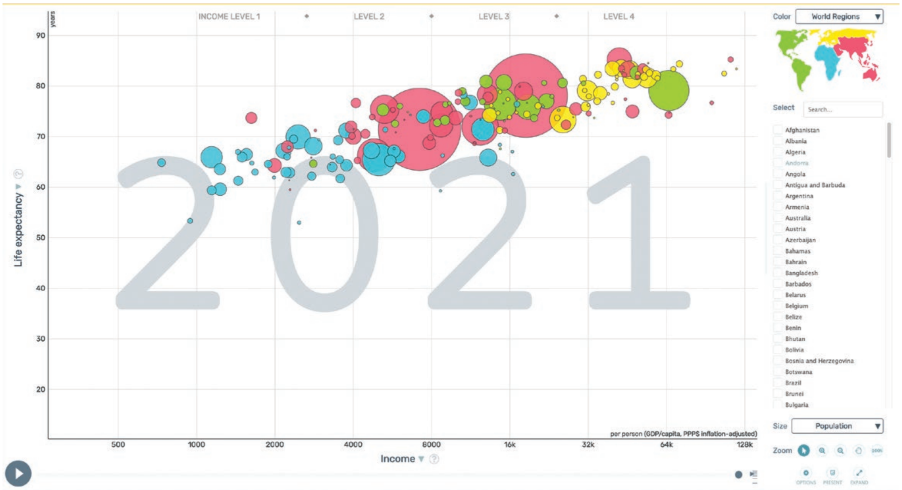

Figure 8.2. Gapminder animated scatterplot showing the relationship between life expectancy and income for people in countries around the world over the last two centuries (Rosling et al., 2021). Free material from www.gapminder.org. CC BY GAPMINDER.ORG.

Overall, then, by visualizing patterns, trends, and outliers, data design enables contemporary audiences to perceive quantitative aspects of the world around them in highly accessible forms. Data visualization is especially valuable in the field of technical communication, where facts and data must be communicated clearly and efficiently to both professional and lay audiences.

References

Barton, B. F., & Barton, M. S. (1993a). Ideology and the map: Toward a postmodern visual design practice. In N. Roundy Blyler & C. Thralls (Eds.), *Professional communication: The social perspective* (pp. 49-78). Sage.

Barton, B. F., & Barton, M. S. (1993b). Modes of power in technical and professional visuals. *Journal of Business and Technical Communication*, *7*(1), 138-162. https://doi.org/10.1177/1050651993007001007

Battle-Baptiste, W., & Rusert, B. (Eds.). (2018). *W.E.B. Du Bois's data portraits: Visualizing Black America: The color line at the turn of the twentieth century*. The W.E.B. Du Bois Center at the University of Massachusetts Amherst; Princeton Architectural Press.

Bertin, J. (1981). *Graphics and graphic information processing*. (W. J. Berg & P. Scott, Trans.). Walter De Gruyter. https://doi.org/10.1515/9783110854688

Brasseur, L. E. (2003). *Visualizing technical information: A cultural critique*. Baywood. https://doi.org/10.2190/VTI

Brinton, W. C. (1914). *Graphic methods for presenting facts*. Engineering Magazine Co.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, *79*(387), 531-554. https://doi.org/10.1080/01621459.1984.10478080

Cochran, J. K., Albrecht, S. A., & Green, Y. A. (1989). Guidelines for evaluating graphical designs: A framework based on human perception skills. *Technical Communication*, *36*(1), 25-32.

Desnoyers, L. (2011). Toward a taxonomy of visuals in science communication. *Technical Communication*, *58*(2), 119-134.

D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. MIT Press. https://doi.org/10.7551/mitpress/11805.001.0001

Dragga, S., & Voss, D. (2001). Cruel pies: The inhumanity of technical illustrations. *Technical Communication*, *48*(3), 265-274.

Eppler, M. J., & Muntwiler, C. (2022). *Interactive knowledge maps*. Visual-literacy.org. https://www.visual-literacy.org/maps/

Ferdio ApS. (2021). *Data Viz Project*. https://datavizproject.com/

Few, S. (2012). *Show me the numbers: Designing tables and graphs to enlighten* (2nd ed.). Analytics Press.

Fisher, D. (2010). Animation for visualization: Opportunities and drawbacks. In J. Steele & N. Iliinsky (Eds.), *Beautiful visualization: Looking at data through the eyes of experts* (pp. 329-352). O'Reilly Media.

Friendly, M. (2008). The golden age of statistical graphics. *Statistical Science*, *23*(4), 502-535. https://doi.org/10.1214/08-STS268

Friendly, M., & Denis, D. J. (2001-2018). *Milestones in the history of thematic cartography, statistical graphics, and data visualization.* http://www.datavis.ca/milestones/

Funkhouser, H. G. (1937). Historical development of the graphical representation of statistical data. *Osiris*, *3*, 269-404. https://doi.org/10.1086/368480

Harris, R. L. (1996). *Information graphics: A comprehensive illustrated reference.* Management Graphics.

Heer, J., & Bostock, M. (2010). Crowdsourcing graphical perception: Using Mechanical Turk to assess visualization design. In *CHI '10: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 203-212). ACI Digital Library. https://doi.org/10.1145/1753326.1753357

Holmes, N. (1984). *Designer's guide to creating charts & diagrams*. Watson-Guptill.

Kienzler, D. S. (1997). Visual ethics. *The Journal of Business Communication*, *34*(2), 171-187. https://doi.org/10.1177/002194369703400204

Kimball, M.A., & Kostelnick, C. (Eds.). (2016). *Visible numbers: Essays on the history of statistical graphics*. Ashgate.

Kostelnick, C. (2004). Melting-pot ideology, modernist aesthetics, and the emergence of graphical conventions: The Statistical Atlases of the United States, 1874-1925. In C. A. Hill & M. Helmers (Eds.), *Defining visual rhetorics* (pp. 215-242). Erlbaum.

Kostelnick, C. (2016a). Mosaics, culture, and rhetorical resiliency: The convoluted genealogy of a data display genre. In M. A. Kimball & C. Kostelnick (Eds.), *Visible numbers: Essays on the history of statistical graphics* (pp. 177-206). Ashgate.

Kostelnick, C. (2016b). The re-emergence of emotional appeals in interactive data visualization. *Technical Communication*, *63*(2), 116-135.

Li, L. (2020). Visualizing Chinese immigrants in the *U.S. Statistical Atlases*: A case study in charting and mapping the other(s). *Technical Communication Quarterly*, *29*(1), 1-17. https://doi.org/10.1080/10572252.2019.1690695

Macdonald-Ross, M. (1977). How numbers are shown: A review of research on the presentation of quantitative data in texts. *AV Communication Review*, *25*(4), 359-409. https://doi.org/10.1007/BF02769746

Neurath, O. (1939). *Modern man in the making*. Alfred A. Knopf.
Playfair, W. (1801). *The commercial and political atlas, representing, by means of stained copper-plate charts, the progress of the commerce, revenues, expenditure, and debts of England, during the whole of the eighteenth century* (3rd ed.). T. Burton.
Rawlins, J. D., & Wilson, G. D. (2014). Agency and interactive data displays: Internet graphics as co-created rhetorical spaces. *Technical Communication Quarterly*, *23*(4), 303-322. https://doi.org/10.1080/10572252.2014.942468
Ribecca, S. (2023). *The data visualisation catalogue*. https://datavizcatalogue.com/index.html
Rosling, O., Rosling Rönnlund, A., & Rosling, H. (2021). *Wealth and health of nations* [Interactive bubble chart]. Gapminder. https://www.gapminder.org/tools/?-from=world#$chart-type=bubbles
Sloane, C. S., & U.S. Bureau of the Census (1914). *Statistical atlas of the United States.* U.S. Government Printing Office.
Tableau [Computer software]. (2023). https://www.tableau.com/
Tufte, E. R. (1983). *The visual display of quantitative information*. Graphics Press.
Tufte, E. R. (1990). *Envisioning information*. Graphics Press.
Walker, F. A., & U.S. Census Office (1874). *Statistical atlas of the United States based on the results of the ninth census 1870*. Julius Bien; Library of Congress Geography and Map Division. https://www.loc.gov/item/05019329/
Yau, N. (2013). *Data points: Visualization that means something*. John Wiley & Sons.